Original Article

# Nonlinear effects of group size on the success of wolves hunting elk

**Daniel R. MacNulty,[a] Douglas W. Smith,[b] L. David Mech,[c] John A. Vucetich,[d] and Craig Packer[a]**
[a]Department of Ecology, Evolution, and Behavior, University of Minnesota, 1987 Upper Buford Circle, St. Paul, MN, 55108, USA, [b]Yellowstone Center for Resources, PO Box 168, Yellowstone National Park, WY, 82190, USA, [c]U.S. Geological Survey, Northern Prairie Wildlife Research Center, 8711 37th Street SE, Jamestown, ND, 58401, USA, and [d]School of Forest Resources and Environmental Science, 1400 Townsend Dr., Houghton, MI, 49931, USA

Despite the popular view that social predators live in groups because group hunting facilitates prey capture, the apparent tendency for hunting success to peak at small group sizes suggests that the formation of large groups is unrelated to prey capture. Few empirical studies, however, have tested for nonlinear relationships between hunting success and group size, and none have demonstrated why success trails off after peaking. Here, we use a unique dataset of observations of individually known wolves (*Canis lupus*) hunting elk (*Cervus elaphus*) in Yellowstone National Park to show that the relationship between success and group size is indeed nonlinear and that individuals withholding effort (free riding) is why success does not increase across large group sizes. Beyond 4 wolves, hunting success leveled off, and individual performance (a measure of effort) decreased for reasons unrelated to interference from inept hunters, individual age, or size. But performance did drop faster among wolves with an incentive to hold back, i.e., nonbreeders with no dependent offspring, those performing dangerous predatory tasks, i.e., grabbing and restraining prey, and those in groups of proficient hunters. These results suggest that decreasing performance was free riding and that was why success leveled off in groups with >4 wolves that had superficially appeared to be cooperating. This is the first direct evidence that nonlinear trends in group hunting success reflect a switch from cooperation to free riding. It also highlights how hunting success per se is unlikely to promote formation and maintenance of large groups. *Key words: Canis lupus*, carnivore, cooperation, free riding, group hunting, group living, interference, predation, sociality, wolf. *[Behav Ecol]*

## INTRODUCTION

Improved ability to capture prey is a well-known benefit of group living in social predators and a classic explanation for the evolution of sociality in group-living predators (Alexander 1974; Kruuk 1975; Pulliam and Caraco 1984; Clark and Mangel 1986). However, it is much less appreciated that the benefit of improved hunting success (defined as the likelihood of capturing prey) may only be realized in small groups. Data from a range of social predators suggest that success initially increases, then levels off, or even declines with group size despite apparent cooperation among hunters (reviewed by Packer and Ruttan 1988; see also Boesch C and Boesch H 1989; Boesch 1994; Rose 1997; Kim et al. 2005). This pattern is exemplified in large social carnivores, which have been the focus of much research on group hunting behavior. Many studies show that carnivore hunting success peaks at 2–5 hunters and remains constant or declines over larger group sizes (Eaton 1970; Kruuk 1972; Schaller 1972; Van Orsdol 1984; Mills 1985; Stander 1992; Fanshawe and Fitzgibbon 1993; Holekamp et al. 1997; Funston et al. 2001). Thus, group hunt-

ing success may play no role in the formation and maintenance of large groups as is often assumed. On the other hand, the actual peak of success at small group sizes is uncertain because studies of group hunting success have seldom tested for nonlinear relationships between success and group size (cf. Holekamp et al. 1997), and none have demonstrated why success trails off after peaking.

There are 2 prevailing hypotheses for why group size–specific hunting success ($H_n$) is nonlinear. The interference hypothesis proposes that $H_n$ is limited in large groups because individual predators impede each other's actions. This predicts that the rate of decline in $H_n$ is greatest when individual hunters are inept because they would be most likely to get in each other's way. There is evidence that inept predators (e.g., juveniles) reduce the magnitude of $H_n$ (Creel S and Creel NM 1995; Funston et al. 2001), but the extent to which they change its slope (i.e., from positive to zero or negative slope) remains untested. It is also possible that simple overcrowding of even very adept predators can reduce success. If so, the latency to a successful hunt should increase with group size given that overcrowding involves time-consuming interactions between predators whether they are adept or inept hunters. This is well demonstrated in foraging experiments involving homogenous groups of robots (Balch and Arkin 1994; Beckers et al. 1994; Lerman and Galstyan 2002).

The free-rider hypothesis states that $H_n$ is limited in large groups because individual predators withhold effort and

---

participate only to remain nearby to gain access to the kill. In this case, "free riding" is the act of withholding contributions toward making a kill (a collective good), while benefiting from the effort of others (Nunn and Lewis 2001). Free riding is expected to occur regardless of whether an individual is an adept or inept hunter. The central prediction is that hunters, irrespective of hunting ability, reduce effort beyond the group size at which $H_n$ levels off or declines. According to theory, hunters start holding back at this group size because it is where the costs of hunting (e.g., risk of injury and energetic loss) exceed the diminishing improvements in group hunting success with each additional hunter (Packer and Ruttan 1988). Scheel and Packer (1991) found that individual lions (*Panthera leo*) free ride in large groups, but they observed too few kills to relate this to changes in $H_n$.

Here, we use a novel dataset derived from direct observations of individually known wolves (*Canis lupus*) hunting elk (*Cervus elaphus*) in Yellowstone National Park to 1) quantify the nonlinear effects of group size on hunting success using modern statistical techniques and 2) test whether nonlinearity in $H_n$ is a consequence of interference from inept hunters or free riding. Given that a successful hunt is the product of success at each of several component phases (Lima and Dill 1990), we evaluate nonlinearity in $H_n$ by measuring the influence of group size on whether a group completes each of 3 predatory tasks (attacking, selecting, and killing) corresponding to the transitions between 4 behaviors (approach, attack-group, attack-individual, and capture; see Table 1 for definitions) that comprise the typical predatory sequence of cursorial carnivores hunting social ungulates (MacNulty et al. 2007). We then examine the behavior of individually known wolves and test how group size affects an individual's performance of each task after controlling for its age and body mass, which are key determinants of wolf hunting ability (MacNulty, Smith, Mech, et al. 2009; MacNulty, Smith, Vucetich, et al. 2009).

If interference from inept hunters limits $H_n$, we expected decreases in the collective ability of groups to reduce success with increasing group size at each phase of the hunt. Conversely, if free riding limits $H_n$, individual performance—a measure of hunting effort—should drop beyond any peak in $H_n$ irrespective of group ability. And if this decline was free riding in response to high hunting costs, we predicted that the rate of decline followed killing > selecting > attacking as this reflects between-task differences in the risk of injury (MacNulty, Smith, Vucetich, et al. 2009). And given that wolves live in cooperatively breeding family groups (Mech and Boitani 2003), we also expected performance to decline more rapidly in nonbreeding wolves than in breeding wolves because nonbreeders lack dependent offspring (i.e., pups), and this should discourage an all-out hunting effort.

**Table 1**
**Ethogram of wolf predatory behavior**

| Foraging State | Definition |
| --- | --- |
| Approach | Fixating on and traveling toward prey. |
| Attack-group | Running after a fleeing prey group or lunging at a standing group while glancing about at different group members (i.e., scanning) |
| Attack-individual | Running after or lunging at a solitary prey or a single member of a prey group although ignoring all other group members. |
| Capture | Biting and restraining prey. |

See MacNulty et al. (2007) for additional details.

## METHODS

### Study area

Yellowstone National Park extends across 891 000 ha of a primarily forested plateau in northwestern Wyoming, USA, that ranges from 1500 to 3300 m. Large montane grasslands provide excellent views of wildlife. Observations of wolves hunting were made primarily in a 100 000 ha grassland complex in the northeastern quarter of Yellowstone referred to as the Northern Range. This area is characterized by a series of open valleys, ridges, and minor plateaus. Low elevations (1500–2000 m) there create the warmest and driest conditions in Yellowstone during winter, providing critical winter range for ungulates, including mainly elk (Houston 1982). A maintained road runs the length of the Northern Range and provides year-round vehicle access.

### Study population

A combined total of 41 radio-marked wolves were reintroduced to Yellowstone National Park in 1995–1997 (Bangs and Fritts 1996). Wolves observed in this study were either members or descendents of the original reintroduced population. In each year after the reintroduction, approximately 30–50% of the pups born were captured and radio-marked (Smith et al. 2004) following applicable animal handling guidelines of the American Society of Mammalogists (Animal Care and Use Committee 1998). In this study, we focused mainly on 5 wolf packs: Druid Peak, Geode Creek, Leopold, Mollie's, and Rose Creek. At least 2 individuals in each pack were radio-marked. A total of 94 wolves were individually identifiable by combination of radio frequency, pelage color, body shape, and/or size. These wolves were the focus of our individual-level hunting analysis (see below) and were observed for 1–8 years (1995–2003).

We annually classified the breeding status (breeder/non-breeder) of each wolf according to whether it whelped or sired pups each spring (April). Breeding wolves included the socially dominant male and female of each pack and occasionally ≥1 subordinate female (vonHoldt et al. 2008). Nonbreeding wolves included mature (≥2 years old) and immature (yearling) offspring from previous litters as well as adults unrelated to the breeders. Only 1 nonbreeder (5F) was ever socially dominant (Mech et al. 1996). Mean (±SE) age (years) of breeders and nonbreeders was $4.87 \pm 0.09$ (range = 1.06–9.64) and $1.86 \pm 0.04$ (range = 0.54–7.85), respectively.

### Behavior sampling

The methods we used to observe and record hunting behavior were described elsewhere (MacNulty et al. 2007), and here, we highlight only key aspects relevant to the current analysis. Various assistants and 2 of the authors (DRM and DWS) observed wolves hunting elk during biannual 30-day follows of 3–14 packs from the ground and fixed-wing aircraft in early (mid-November to mid-December) and late (March) winter and during opportunistic surveys throughout the remainder of the year (Smith et al. 2004). Wolves hunted mainly elk (MacNulty et al. 2007), and 97% of 469 wolf–elk encounters used in this study were directly observed from the ground in the Northern Range. Most encounters (84%) involved groups of elk.

When wolves encountered elk—defined as at least 1 wolf orienting and moving (walking, trotting, or running) toward elk—we followed the progress of the encounter by noting the foraging state (approach, watch, attack-group, attack-individual, and capture) of the individual(s) closest to making a kill. We therefore recorded the sequential occurrence of the most

escalated state and the number and identity of wolves participating in that state. A wolf was scored as participating in a foraging state if it exhibited the behavioral acts characterizing that particular state (Table 1). We considered nonparticipation in a given state as when a wolf was in view but engaged in another foraging state or nonpredatory behavior, e.g., resting. We refer to the wolves participating in a foraging state as the "hunting group".

We scored group hunting success and individual performance according to whether wolves completed each of 3 predatory tasks that corresponded to the following 3 behavioral transitions: approach → attack-group = "attacking"; attack-group → attack-individual = "selecting"; and attack-individual → capture = "killing". If an individual wolf participated in a pair of consecutive foraging states that comprised a given task, it was scored as having performed that task. Nonperformance was when an individual failed to complete a task-specific transition (e.g., attack-group→ approach). A hunting group completed a task, and was therefore "successful", if the task was performed by at least 1 group member. If not, we considered the group to have "failed" in that task. This scheme generated a binary score for a group and each of its members in each sequential foraging state.

**Data analysis**

To assess the effects of group size on hunting success, and the relationship between group-level success and individual-level performance, we analyzed how hunting group size (i.e., number of wolves participating in a foraging state) influenced the probability that groups and individuals attacked, selected, and killed elk based on the binary scores described above. This involved a separate group- and individual-level analysis in which we separately analyzed the effects of hunting group size on the completion of each predatory task. We limited our analyses of killing to adult elk to control for the effects of prey size on group hunting behavior (Packer and Ruttan 1988). Analyses were performed with generalized linear mixed models (GLMMs) with a binomial error distribution. Such models can account for correlation between the multiple observations taken on each identifiable wolf ($N = 94$) and each pack ($N = 5$). Individual and pack identity were fitted as a random effect in the individual- and group-level models, respectively. Note that the random effect for pack identity accounts for the influence of unmeasured pack-related factors on hunting success, including differences in prey density between pack territories.

Observations of repeated attempts to perform the same task during the same encounter were also correlated, but these were used in only the group-level models of selecting and killing, which fitted encounter identity as a random effect within pack. Models of attacking included only the first attempt because we were mainly interested in how group size affected the probability of attack on first encountering elk. Similarly, the individual-level models included only the first attempts to select and kill because these models had trouble converging when we included >1 random effect. All models included a compound symmetric correlation structure, which assumed that all observations within packs, individuals, or encounters were, on average, equally correlated (Weiss 2005). Models were estimated with adaptive Gaussian quadrature, with parameters estimated from maximum likelihood, and significance of effects determined by an approximate z-test.

We used piecewise linear splines to test for nonlinear effects of group size on the probability that groups and individuals completed a given predatory task. Piecewise splines consist of a continuous covariate (e.g., group size) defined over specified segments (e.g., > and <4 wolves) and a response variable

(e.g., hunting success) that is a continuous function of the covariate over all segments but with different slopes in each of the segments (Marsh and Cormier 2002). Each line segment does not have its own intercept. Rather, a spline regression model includes only a single intercept that is adjusted by the spline variable to accommodate a change in slope. This keeps the regression line constant (i.e., no breaks) even as the regression line pivots to change direction at the point(s) where the segments join. In the spline literature, the join points are called knots. In epidemiology, knot location is used to identify the threshold value of a risk factor for which the probability of disease occurring suddenly changes (Bessaoud et al. 2005). Similarly, we used splines to identify the threshold group size beyond which the probability of group hunting success and individual performance abruptly changes.

Splines are an improvement over low dimension polynomials (e.g., quadratics) because they allow sudden changes in slope at irregular intervals. By contrast, the fit of a polynomial curve over one region of the data is directly affected by the fit of the curve elsewhere. And in model estimation, the coefficients for spline variables are usefully interpreted as either the change in slope between line segments or the slopes of the line segments themselves, whereas polynomial coefficients have no equivalent interpretation (Eubanks 1984; Marsh and Cormier 2002; MacKenzie et al. 2005).

To determine the presence and position of group size–specific thresholds in task performance, we evaluated a set of competing GLMMs for each task. Each model set included models with a single knot placed at 2–9 hunters, a model with no knot representing the hypothesis of no threshold in task performance, and an intercept-only model representing the null hypothesis that group size had no affect on hunting success. We selected knots a priori according to previously cited reports that carnivore hunting success peaks in groups with <10 hunters. Our placement of knots is consistent with guidelines for the efficient use of knots (Wold 1974; Eubanks 1984; Seber and Wild 2003). By definition, knots selected a priori are fixed (i.e., not random variables) and are therefore not estimated as parameters in models. We created variables containing a linear spline for group size with the MKSPLINE command in STATA 10.1. The variables were constructed so that the estimated coefficients measure the slopes for the segments before and after a given knot.

All candidate models in the individual-level analysis included terms for age and body mass to control for age- and size-specific variation in hunting ability. Body mass affects wolf predatory performance independently of age and also accounts for the main effect of sex on performance (MacNulty, Smith, Mech, et al. 2009). Mass was estimated from an individually based sex-specific growth model derived from measurements of 304 wolves, including 86 focal wolves (see MacNulty, Smith, Mech, et al. 2009 for details). Details about how we estimated age were given elsewhere (MacNulty, Smith, Vucetich, et al. 2009).

To determine if interference explained the relationship between hunting success and group size, we tested the hypothesis that inept hunters were responsible for a decline in success in larger groups. To do so, we analyzed the interaction between group size and group ability for a subset of observations ($N = 88–147$ wolf–elk encounters) in which the identity of each group member was known. A wolf's identity provided information on its age, sex, and body mass, which we used to estimate its probability of performing a given task based on previous analyses of wolf hunting ability (MacNulty, Smith, Mech, et al. 2009; MacNulty, Smith, Vucetich, et al. 2009). The predicted probability of performing was based on a linear predictor that included both fixed and random effects and so was conditional on the values of the estimated random effects. Next, each group member was ranked from 1 (worse) to 10

(best) according to its expected ability to perform. Ranks were determined for each task by applying *k*-means cluster analysis to the conditional probabilities of performance calculated from the models and data (*N* = 189–281 wolf–elk encounters) presented by MacNulty, Smith, Vucetich, et al. (2009). These models excluded group size because it had no effect on age- or size-specific variation in hunting ability. We used the median of the ranks of each group member as a measure of a group's relative hunting ability and evaluated whether it altered the relationship between hunting success and group size by testing whether interactions between median group rank and group size improved the fit of the top group-level and individual-level models. We also assessed interference in terms of how group size influenced the time to complete each task.

We conducted all analyses in STATA 10.1 and compared GLMMs using information-theoretic statistics (Burnham and Anderson 2002). Our scope of inference concerned the population, so we performed model selection using marginal likelihoods. The most parsimonious model was the one with the lowest Akaike Information Criterion (adjusted for small sample, $AIC_c$) and smallest $\Delta AIC_c$. $\Delta AIC_c$ equals the $AIC_c$ for the model of interest minus the smallest $AIC_c$ for the set of models being considered. The best model has a $\Delta AIC_c$ of zero, and

models with $\Delta AIC_c$ <2 are plausibly the best. To assess uncertainty about the best model, we identified models with $\Delta AIC_c$ <2 as the confidence set of models (analogous to a confidence interval for a mean estimate; Burnham and Anderson 2002). We calculated population-averaged fitted values from best-fit GLMMs by deriving marginal expectations of the responses averaged over the random effects but conditional on the observed covariates. We also used likelihood ratio statistics to test specific hypotheses among nested models, and results were considered significant at $P < 0.05$. Means are reported with standard errors unless indicated otherwise.

## RESULTS

### Group hunting success

The influence of group size on wolf hunting success was nonlinear (Figure 1a–c). The most parsimonious models of attacking, selecting, and killing include a linear spline for group size (see Supplementary Table S1), indicating a threshold at which the affect of group size on hunting success suddenly changed. Evidence against a model describing a simple linear relationship between group size and success is reasonably strong for attacking ($\Delta AIC_c$ = 5.93) and killing ($\Delta AIC_c$ = 2.46) but somewhat



**Figure 1**
Main effects of hunting group size on the probability that wolf packs attack (a), select (b), and kill (c) elk. Open circles are population-averaged fitted values with 95% confidence intervals from the best-fit GLMM models of pack-level hunting success (Supplementary Table S1). The estimated coefficients before and after each breakpoint are $0.37 \pm 0.12$ ($P = 0.002$) and $-0.05 \pm 0.05$ ($P = 0.281$) (a); $0.36 \pm 0.16$ ($P = 0.023$) and $0.04 \pm 0.05$ ($P = 0.467$) (b); $0.36 \pm 0.19$ ($P = 0.053$) and $-0.21 \pm 0.13$ ($P = 0.124$) (c). The number of wolf–elk encounters included in each analysis is 355 (a), 376 (b), and 235 (c). Filled circles are observed frequencies with sample size indicated above each point. Analyses were performed on the raw binary data and not the illustrated data points, which are provided as a visual aid. The product of the fitted value lines and associated confidence intervals in (a), (b), and (c), representing the overall probability of success given an elk encounter and thus the net effect of group size on group hunting success, is shown in (d).

weak for selecting ($\Delta AIC_c = 1.12$; see Supplementary Table S1). Thus, we are less certain about a threshold in selecting than in attacking and killing. Moreover, the intercept model fits the killing data surprisingly well ($\Delta AIC_c = 0.42$; see Supplementary Table S1c), implying that the overall influence of group size on killing was not strong. This was not the case for attacking or selecting (intercept model: $\Delta AIC_c > 5.36$; see Supplementary Tables S1a,b).

The threshold group size was relatively small. The confidence set of spline models ($\Delta AIC_c < 2$) for each predatory task (see Supplementary Table S1) indicates the threshold group size was 4–7 wolves for attacking and 2–6 wolves for selecting and killing. The most parsimonious models in the set include a threshold at 4 wolves for attacking and killing and a threshold at 3 wolves for selecting (Figure 1a–c). The product of these models' population-averaged fitted values and associated pointwise 95% confidence intervals, which represents the net effect of group size across all tasks (sensu MacNulty, Smith, Mech, et al. 2009), reveals that overall hunting success [$P(\text{kill}|\text{encounter})$] peaked at 4 wolves (Figure 1d). Note that multiplying confidence intervals across tasks probably exaggerates variability in overall success due to positive correlations between tasks. Thus, confidence intervals in Figure 1d are conservative.

Importantly, hunting success did not measurably improve beyond 3–4 wolves. According to the best-fit models, group size had no significant effect on success once it exceeded each task-specific threshold ($P = 0.12$–0.48; Figure 1). Below these thresholds, each additional wolf improved group hunting success by 45% (odds ratio [OR] $= 1.45 \pm 0.17$, $P = 0.002$), 43% (OR $= 1.43 \pm 0.23$, $P = 0.023$), and 44% (OR $= 1.44 \pm 0.72$, $P = 0.053$) in attacking, selecting, and killing, respectively. Results were the same for a subset of observations that included data on elk group size; success leveled off at 3–4 hunters regardless of elk group size.

The asymptote in group hunting success was apparently unrelated to interference from inept hunters or delays in task completion. Likelihood ratio $\chi^2$ values and associated $P$ values (from nested-model comparisons) are nonsignificant for interactions between group size and group ability beyond each task-specific threshold (attacking: $\chi_1^2 = 0.90$, $P = 0.34$; selecting: $\chi_1^2 = 0.01$, $P = 0.94$; and killing: $\chi_1^2 = 0.12$, $P = 0.73$). Group ability was generally high: for each task, the median group hunting rank was 7 (interquartile ranges $= 2.75$–4.00). Data on time to accomplish each task are best fit by an intercept model (see Supplementary Table S2), indicating no effect of group size on latency to a successful hunt. For each task, a simple linear model describing an increase in latency with group size provides the next best fit to the data ($\Delta AIC_c$ 0.27–1.98; see Supplementary Table S2); yet, in each case, the coefficient for group size is not significant ($P = 0.16$–0.75).
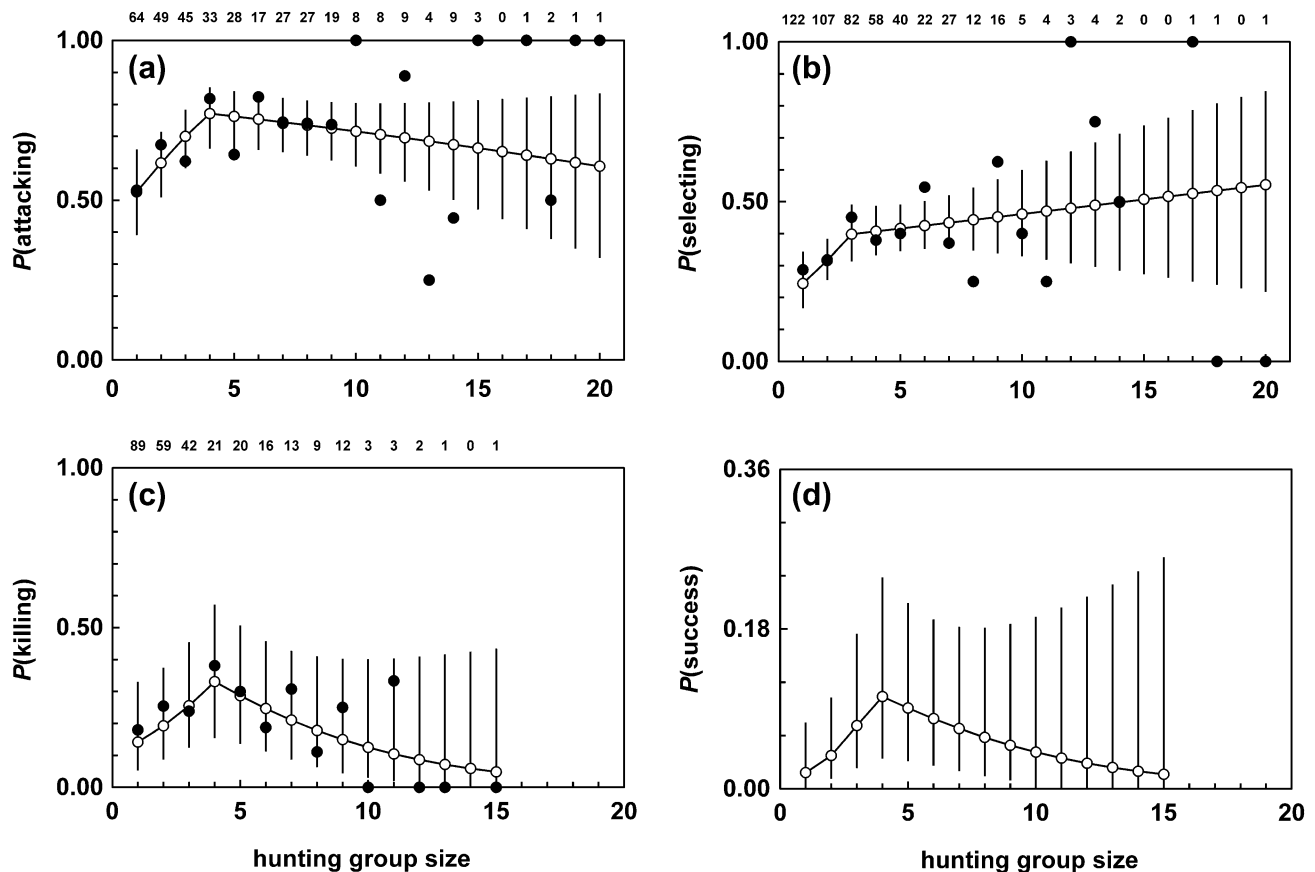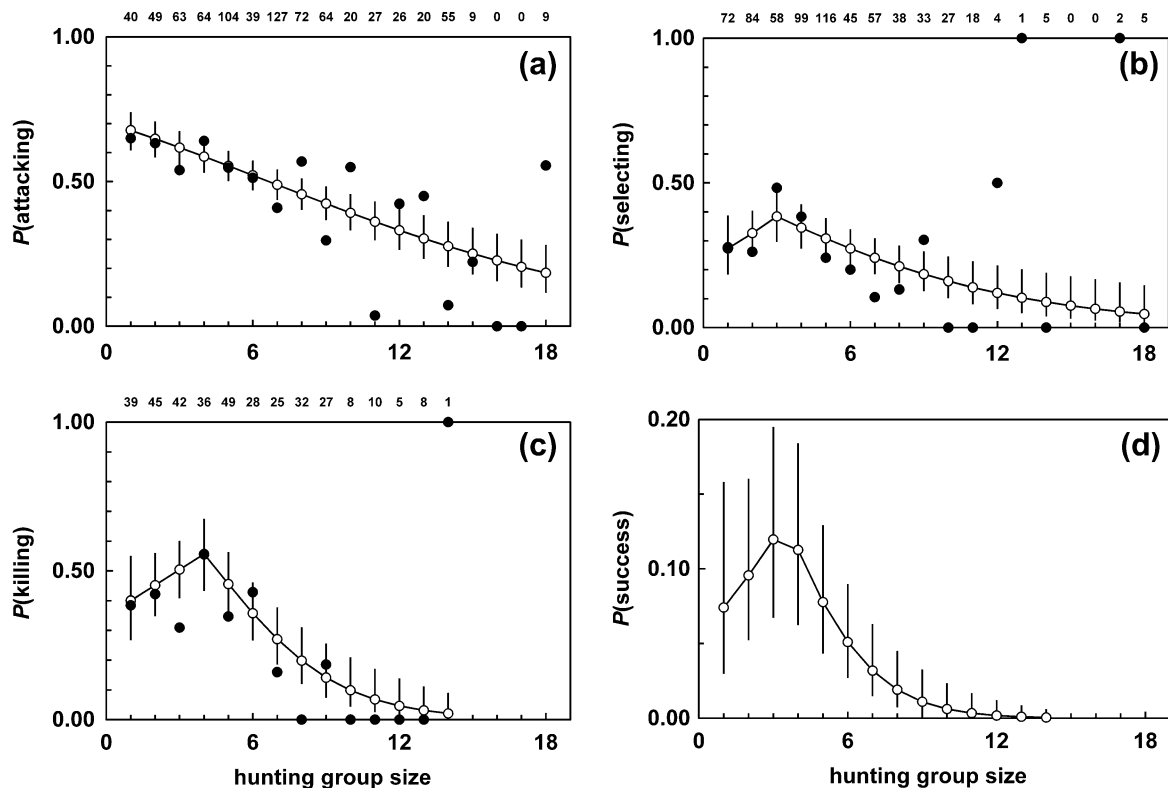


**Figure 2**
Main effects of hunting group size on the probability that individual wolves attack (a), select (b), and kill (c) elk. Open circles are population-averaged fitted values with 95% confidence intervals from the best "group size" GLMM models of individual-level predatory performance (Supplementary Table S3a-S5a). The estimated coefficients are $-0.13 \pm 0.02$ ($P < 0.001$) (a); $0.25 \pm 0.16$ ($P = 0.121$) and $-0.17 \pm 0.05$ ($P = 0.001$) (b); $0.21 \pm 0.13$ ($P = 0.115$) and $-0.41 \pm 0.09$ ($P < 0.001$) (c). The number of wolves and wolf–elk encounters included in each analysis follows: 86 and 254 (a); 81 and 278 (b); 70 and 153 (c). Filled circles are observed frequencies with sample size indicated above each point. Analyses were performed on the raw binary data and not the illustrated data points, which are provided as a visual aid. The product of the fitted value lines and associated confidence intervals in (a), (b), and (c), representing the overall probability that an individual kills an elk given an encounter and thus the net effect of group size on individual-level predatory performance, is shown in (d).

## Individual performance

We first constructed a set of "group size" models in which only group size (i.e., subject + companions) describes individual performance after controlling for individual age and body mass. Overall, this analysis reveals that individual performance decreased with increasing group size, and it did so at or near the group size at which group hunting success leveled off. For instance, an individual wolf was decreasingly likely to select and kill an elk beyond 3 and 4 wolves, respectively (Figure 2b,c), which match the inferred peaks in group-level selecting and killing (Figure 1b,c). Evidence against alternative individual-level models of selecting and killing is reasonably strong ($\Delta AIC_c > 2.00$; see Supplementary Table S4a and S5a), except for a model of killing with a threshold at 5 wolves ($\Delta AIC_c = 1.35$; see Supplementary Table S5a). And although a simple linear model fits the individual attacking data best (Figure 2a), the next best model includes a threshold at 4 wolves ($\Delta AIC_c = 1.35$; see Supplementary Table S3a), which matches the threshold in group-attacking success (Figure 1a). The peak in overall individual performance (Figure 2d) also corresponds well with the peak in overall group hunting success (Figure 1d). In general, these results indicate that decreasing individual performance was responsible for the lack of improvement in group hunting success in groups with >4 wolves.

That the decrease in individual performance was the result of wolves withholding hunting effort in response to hunting costs is evidenced by between-task differences in the rate of declining performance. Specifically, the rate of decline was fastest for the most dangerous task (killing: $-0.41 \pm 0.09$, $P < 0.001$; Figure 2c) and slowest for the safest task (attacking: $-0.13 \pm 0.02$, $P < 0.001$; Figure 2a). Combining task-specific datasets and testing for interactions between task type and group size beyond >3–4 wolves revealed that the rate of decline was significantly faster for killing than for attacking ($z = -3.14$, $P = 0.002$) or selecting ($z = -2.70$, $P = 0.007$). The decline in selecting ($-0.17 \pm 0.05$, $P = 0.001$; Figure 2b) was faster than for attacking, but the difference was not statistically significant ($z = -1.26$, $P = 0.21$).

To determine whether individual breeding status (i.e., breeder/nonbreeder) or age influenced the relationship between group size and individual performance, we tested whether the effect of group size varies according to an individual's breeding status or age. To do so, we added interaction terms (group size × breeder, group size × age) to the "group size" models (see Supplementary Tables S3a–S5a) to produce a set of "breeder-varying" (see Supplementary Tables S3b–S5b) and "age-varying" models (see Supplementary Tables S3c–S5c). Next, we compared models across all 3 sets for each task. The best overall models of attacking and killing include a positive group size × breeder interaction but are otherwise identical to those identified in the "group size" set (see Supplementary Tables S3b–S5b). These "breeder-varying" models fit the data better than their "group size" analogs (attacking: $\chi_2^2 = 14.07$, $P < 0.001$ and killing: $\chi_2^2 = 7.58$, $P = 0.023$) and indicate that a breeding wolf was more likely to attack and kill in large groups than was a nonbreeder. A similar breeder-varying model of selecting scored well ($\Delta AIC_c = 0.38$; see Supplementary Table S4b), but it was not significantly different from its group size analog ($\chi_2^2 = 4.76$, $P = 0.093$). None of the age-varying models scored well (attacking: $\Delta AIC_c > 7.40$, see Supplementary Table S3c; selecting: $\Delta AIC_c > 3.30$, see Supplementary Table S4c; and killing: $\Delta AIC_c > 3.50$, see Supplementary Table S5c), indicating that the drop in performance with increasing group size was independent of an individual's age.

Interference from inept hunters had little, if any, influence on declines in individual performance with group size. For a subset of observations in which we have information on group ability, we compared the best overall models of attacking, selecting, and killing (see Supplementary Tables S3–S5) with similar models that include a group size × group ability interaction. Although this interaction tends to improve model fit (attacking: $\chi_1^2 = 4.17$, $P = 0.041$, N individuals/encounters = 48/83; selecting: $\chi_1^2 = 3.61$, $P = 0.057$, $N = 50/92$; and killing: $\chi_1^2 = 3.76$, $P = 0.052$, $N = 34/45$), it is significantly negative in 2 tasks (attacking: $\beta = -0.05 \pm 0.03$, $z = -2.03$, $P = 0.042$ and selecting: $\beta = -0.08 \pm 0.04$, $z = -2.02$, $P = 0.044$), indicating that individuals in incompetent groups were more likely to perform in large groups than were individuals in competent groups. By contrast, the interference hypothesis predicts that individuals are less likely to perform as a consequence of inept companions. For killing, however, the group size × group ability interaction was positive, though not significantly so ($\beta = 0.45 \pm 0.26$, $z = 1.76$, $P = 0.079$), yielding only marginal evidence that individual performance suffered in large groups due to interference from inept hunters.

Finally, we checked if underperformance was due to individuals alternating hunting effort within or between hunts. Conceivably, individuals underperformed because they were exhausted from performing earlier in the same hunt or in a different hunt the same day. These data were sparse, so we pooled daily observations and scored preceding performance for each task according to whether an individual had performed the task at any time earlier in the same day. We tested this variable using our best-fit models of task performance (see Supplementary Tables 3–5) for a subset of observations that included information on preceding performance. Contrary to the alternating effort hypothesis, a wolf was more likely to perform a task if it had performed it earlier that same day (attacking: OR = $3.27 \pm 1.37$, $P = 0.005$, $N = 44/49$; selecting: OR = $2.52 \pm 1.17$, $P = 0.047$, $N = 30/67$; and killing: OR = $6.79 \pm 7.20$, $P = 0.071$, $N = 16/27$).

## DISCUSSION

Contrary to the popular view that increasing group size usually improves hunting success (e.g., Creel S and Creel NM 2002; Sand et al. 2006), wolves hunting elk in Yellowstone National Park did not perform better in groups with >4 wolves. Our data are consistent with results from many other group hunting predators including insects, birds, primates, and other carnivores (Packer and Ruttan 1988; Boesch C and Boesch H 1989; Boesch 1994; Rose 1997; Kim et al. 2005). In most carnivore studies, for example, hunting success appears to level off beyond 2–5 hunters (Eaton 1970; Kruuk 1972; Schaller 1972; Van Orsdol 1984; Mills 1985; Stander 1992; Fanshawe and Fitzgibbon 1993; Holekamp et al. 1997; Funston et al. 2001). But because so few empirical studies have actually tested for nonlinear relationships between group size and hunting success (cf. Holekamp et al. 1997), large groups are often assumed to be more successful than groups of 2–5 individuals.

Our results provide unique empirical insight into the behavioral mechanisms that prevent increased hunting success in larger groups. Assessing these mechanisms is challenging because it requires repeated observations of known individuals, complete knowledge of group composition, and frequent observation of successful hunts. As a result, previous analyses of the effects of inept hunters on $H_n$ (group size–specific hunting success) are limited to measuring impacts of juvenile hunters (Creel S and Creel NM 1995; Funston et al. 2001), which are easily identified by virtue of their smaller sizes. Yet, variation in hunting ability is known to include substantial variation among adults; in some cases, adults are even less capable than juveniles owing to senescence (MacNulty, Smith, Vucetich

et al. 2009). This finding motivated our ranking of individual hunting ability according to age-specific models of performance that control for age-related declines (MacNulty, Smith, Vucetich et al. 2009).

Importantly, the number of inept hunters in a wolf hunting group did not explain why $H_n$ decreased across large groups: the rate of decline in $H_n$ was independent of our measure of group hunting competence. This is an important result because interference from inept hunters is a classic, though untested, hypothesis for why $H_n$ decreases in large groups (Packer and Ruttan 1988).

Could interference between competent hunters have limited $H_n$ at large group size? Such an effect is evident in foraging experiments involving homogenous groups of robots, where group task efficiency (i.e., time to task completion) is reduced in groups with >4 robots due to increases in the number of time-consuming collisions (Balch and Arkin 1994; Beckers et al. 1994; Lerman and Galstyan 2002).

Although robots and wolves show a notable correspondence in peak performance in groups of 4 individuals, we detected no effect of wolf group size on the time to complete a predatory task. This was true even for killing, where competition for space was most likely given that it involved a single elk. Collisions between wolves were probably less likely when attacking and selecting elk than during killing; yet, the relationship between success and group size was similar across all 3 tasks. Thus, none of our results are consistent with the hypothesis that $H_n$ is limited by interference in large groups.

We were able to provide support for the hypothesis that free riding limits $H_n$ in large groups. Consistent with the hypothesis that $H_n$ is limited in large groups due to individuals withholding effort, we found that $H_n$ peaked at the group size where individual performance dropped off. Several lines of evidence suggest that decreasing individual performance resulted from declining effort in response to high hunting costs. First, an individual's performance decreased with increasing group size regardless of its age or the hunting ability of its companions. Thus, we cannot attribute declining performance to an individual's own incompetence or to the collective incompetence of its hunting group.

Second, the rate at which an individual's performance decreased for a given task was correlated with the danger associated with that task. The risk of injury from being kicked, trampled, or stabbed with antlers increases as wolves transition from attacking to selecting to killing and proximity to the elk increases (MacNulty, Smith, Vucetich, et al. 2009). Because of the high fitness costs from injury, the incentive to withhold effort also increases as wolves transition from attacking to selecting to killing. Thus, the more rapid decline in individual performance in large groups through the transition from attacking to selecting to killing (Figure 2a–c), suggests that wolves withhold hunting effort in order to reduce or minimize hunting costs.

Third, the rate at which an individual's performance declined with increasing group size was related to whether it had dependent offspring. If the benefit from provisioning-dependent offspring often exceeds the cost of hunting, breeding members should be less likely to withhold hunting effort than nonbreeding members. Indeed, individual performance in large groups decreased more slowly for breeders than for nonbreeders irrespective of age or weight. This agrees with general findings that breeders usually lead during hunting (Mech and Boitani 2003).

Finally, individual performance when attacking and selecting decreased faster with increasing group size in competent groups than in inept groups. This contradicts the interference hypothesis but supports the free-riding hypothesis insofar as individuals are expected to withhold hunting effort in the presence of competent companions who are likely to succeed by themselves (Packer and Ruttan 1988).

## CONCLUSIONS

Our study suggests that wolves in large groups (>4 hunters) withheld effort, thereby capping further increases in group size–specific hunting success; these individuals likely participated merely to be at hand when a kill was made. A similar increase in free riding with increasing group size is evident in dogs (*Canis lupus familiaris*) involved in intergroup contests (Bonanni et al. 2010). This is also apparent in birds, where scrounging—which is analogous to free riding —increases with group size (Coolen 2002). Cooperation might be more evident when wolves hunt larger prey, as seen in African lions (Scheel and Packer 1991). For example, wolves might cooperate more consistently when hunting bison—which are larger and more formidable than elk (Smith et al. 2000; MacNulty et al. 2007)—if solo hunting success is sufficiently low to leave ample scope for improvement through cooperation (Packer and Ruttan 1988).

That $H_n$ apparently failed to improve owing to increased levels of free riding is consistent with the premise that nonlinear trends in $H_n$ reflect a switch from cooperation to free riding with increasing group size (Packer and Ruttan 1988). Regardless of the mechanism(s) involved, the widespread tendency for hunting success to level off with increasing group size suggests that the influence of group size on hunting success per se is unlikely to promote the formation and maintenance of large predator groups.

## SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://www.beheco.oxfordjournals.org/.

## REFERENCES

Alexander RD. 1974. The evolution of social behavior. Annu Rev Ecol Syst. 5:325–383.

Animal Care and Use Committee. 1998. Guidelines for the capture, handling, and care of mammals as approved by the American Society of Mammalogists. J Mammal. 79:1416–1431.

Balch T, Arkin RC. 1994. Communication in reactive multiagent robotic systems. Auton Robot. 1:27–52.

Bangs EE, Fritts SH. 1996. Reintroducing the gray wolf to central Idaho and Yellowstone National Park. Wildlife Soc B. 24:402–413.

Beckers R, Holland OE, Deneubourg JL. 1994. From local actions to global tasks: stigmergy and collective robotics. Proceedings of the 4th International Workshop on the Synthesis and Simulation of Living Systems Artificial Life. Cambridge (MA): MIT Press. p. 181–189.

Bessaoud F, Daures JP, Molinari N. 2005. Free knot splines for logistic models and threshold selection. Comput Methods Programs Biomed. 77:1–9.

Boesch C. 1994. Cooperative hunting in wild chimpanzees. Anim Behav. 48:653–667.

Boesch C, Boesch H. 1989. Hunting behavior of wild chimpanzees in the Tai National Park. Am J Phys Anthropol. 78:547–573.

Bonanni R, Valsecchi P, Natoli E. 2010. Pattern of individual participation and cheating in conflicts between groups of free-ranging dogs. Anim Behav. 79:957–968.

Burnham KP, Anderson DR. 2002. Model selection and multimodal inference. a practical information-theoretic approach. 2nd ed. New York: Springer-Verlag.

Clark CW, Mangel M. 1986. The evolutionary advantages of group foraging. Theor Pop Biol. 30:45–75.

Coolen I. 2002. Increasing foraging group size increases scrounger use and reduces searching efficiency in nutmeg mannikins (*Lonchura punctulata*). Behav Ecol Sociobiol. 52:232–238.

Creel S, Creel NM. 1995. Communal hunting and pack size in African wild dogs, *Lycaon pictus*. Anim Behav. 50:1325–1339.

Creel S, NM. 2002. The African wild dog: behavior, ecology, and conservation. (NJ): Princeton University Press.

Eaton RL. 1970. Hunting behavior of cheetah. J Wildlife Manage. 34:56–67.

Eubanks RL. 1984. Approximate regression models and splines. Commun Stat Theor M. 13:433–484.

Fanshawe JH, Fitzgibbon CD. 1993. Factors influencing the hunting success of an African wild dog pack. Anim Behav. 45:479–490.

Funston PJ, Mills MGL, Biggs HC. 2001. Factors affecting the hunting success of male and female lions in the Kruger National Park. J Zool. 253:419–431.

Holekamp KE, Smale L, Cooper SM. 1997. Hunting rates and hunting success in the spotted hyena (*Crocuta crocuta*). J Zool. 242:1–15.

Houston DB. 1982. The northern Yellowstone elk: ecology and management. New York: MacMillan.

Kim KW, Krafft B, Choe JC. 2005. Cooperative prey capture by young subsocial spiders I. Functional value. Behav Ecol Sociobiol. 59:92–100.

Kruuk H. 1972. The spotted hyena. Chicago (IL): University of Chicago Press.

Kruuk H. 1975. Functional aspects of social hunting by carnivores. In: Sibly RM, Smith RH, editors. Function and evolution in behavior. Oxford: Blackwell Scientific Publications. p. 521–526.

Lerman K, Galstyan A. 2002. Mathematical model of foraging in a group of robots: effect of interference. Auton Robot. 13:127–141.

Lima SL, Dill LM. 1990. Behavioral decisions made under the risk of predation: a review and prospectus. Can J Zool. 68:619–640.

MacKenzie ML, Donovan CR, McArdle BH. 2005. Regression spline mixed models: a forestry example. J Agric Biol Environ Stat. 10:394–410.

MacNulty DR, Mech LD, Smith DW. 2007. A proposed ethogram of large-carnivore predatory behavior, exemplified by the wolf. J Mammal. 88:595–605.

MacNulty DR, Smith DW, Mech LD, Eberly LE. 2009. Body size and predatory performance in wolves: is bigger better? J Anim Ecol. 78:532–539.

MacNulty DR, Smith DW, Vucetich JA, Mech LD, Stahler DR, Packer C. 2009. Predatory senescence in ageing wolves. Ecol Lett. 12: 1347–1356.

Marsh LC, Cormier DR. 2002. Spline regression models (quantitative applications in the social sciences). Thousand Oaks (CA): Sage.

Mech LD, Boitani L. 2003. Wolf social ecology. In: Mech LD, Boitani L, editors. Wolves: behavior, ecology, and conservation. Chicago (IL): University of Chicago Press. p. 1–34.

Mech LD, Phillips MK, Smith DW, Kreeger TJ. 1996. Denning behavior of non-gravid wolves, *Canis lupus*. Can Field Natur. 110: 343–345.

Mills MGL. 1985. Related spotted hyaenas forage together but do not cooperate in rearing young. Nature. 316:61–62.

Nunn CL, Lewis RJ. 2001. Cooperation and collective action in animal behavior. In: Noe R, van Hoof JARAM, Hammerstein P, editors. Economics in nature. Cambridge: Cambridge University Press. p. 42–66.

Packer C, Ruttan L. 1988. The evolution of cooperative hunting. Am Nat. 132:159–198.

Pulliam HR, Caraco T. 1984. Living in groups: is there an optimal group size? In: Krebs JR, Davies NB, editors. Behavioral ecology: an evolutionary approach. 2nd ed. Sunderland (MA): Sinauer. p. 122–127.

Rose LM. 1997. Vertebrate predation and food-sharing in Cebus and Pan. Int J Primatol. 18:727–765.

Sand H, Wikenros C, Wabakken P, Liberg O. 2006. Effects of hunting group size, snow depth and age on the success of wolves hunting moose. Anim Behav. 72:781–789.

Schaller GB. 1972. The Serengeti lion: a study of predator-prey relations. Chicago (IL): University of Chicago Press.

Scheel D, Packer C. 1991. Group hunting behavior of lions: a search for cooperation. Anim Behav. 41:697–709.

Seber GAF, Wild CJ. 2003. Nonlinear regression. New York: John Wiley and Sons.

Smith DW, Drummer TD, Murphy KM, Guernsey DS, Evans SB. 2004. Winter prey selection and estimation of kill rates in Yellowstone National Park, 1995-2000. J Wildlife Manage. 68:153–166.

Smith DW, Mech LD, Meagher M, Clark WE, Jaffe R, Phillips MK, Mack JA. 2000. Wolf-bison interactions in Yellowstone National Park. J Mammal. 81:1128–1135.

Stander PE. 1992. Foraging dynamics of lions in a semi-arid environment. Can J Zool. 70:8–21.

Van Orsdol KG. 1984. Foraging behavior and hunting success of lions in Queen Elizabeth National Park, Uganda. Afr J Ecol. 22:79–99.

vonHoldt BM, Stahler DR, Smith DW, Earl DA, Pollinger JP, Wayne RK. 2008. The genealogy and genetic viability of reintroduced Yellowstone grey wolves. Mol Ecol. 17:252–274.

Weiss RE. 2005. Modeling longitudinal data. New York: Springer.

Wold S. 1974. Spline functions in data analysis. Technometrics. 16:1–11.